Presentation

Data mining frameworks

- PGMs
- AMIDST
- Stationary data sets
  - Weka
  - R Libs
  - Matlab
- Data streams
  - Elvira
  - Infer.net
  - Hugin
  - Apache SAMOA
  - Vowpal Wabbit

Academic and Industrial partners

- NTNU
- UNI \(\text{DE}\)
- Hugin Expert A/S

Description

- Analysis of big data streams: A complete collection of algorithms for inference and learning of both static and dynamic Bayesian networks from streaming data. Existing software systems for PGMs only focus on stationary datasets.

- Distributed parallel algorithms: AMIDST provides parallel multi-core and distributed implementations of Bayesian parameter learning, using streaming variational Bayes and variational message passing.

Main Features

- Java 8 based
- Latent variable models
- Integration

- Big Data
- Modularity
- Open source

Code example

Learn hidden naive Bayes model from stream

```java
// We can activate the output
parameterLearningAlgorithm.setWindowsSize(100);
// We fix the size of the window
parameterLearningAlgorithm.setWindowsSize(100);
// We can activate the output
parameterLearningAlgorithm.setWindowsSize(100);
```

Use-case: Risk prediction in credit operations

- Concept drift
- Correlated with Unemployment Rate

And much more...

amidst.eu

amidst.github.io/toolbox/